# Charles Creek Flood Zone Modeling: A Correlation Study of Environmental Conditions Versus Water Level in the Pasquotank Watershed

Nartezya Dykes
Computer Science
Spelman College
Atlanta, USA
ndykes@scmail.spelman.edu

Lekedrick Easley
Computer Science
Mississippi Valley State
University
Itta Bena, USA
lekedrickeasley@yahoo.com

Je'aime Powell
Applied Mathematics/
Remote Sensing
Elizabeth City State
University
Bloomington, USA
jeaimehp@gmail.com

Jerome E. Mitchell
Computer Science
Indiana University
Bloomington, USA
jemitchell@cresis.ku.edu

*Abstract*— **The Charles Creek area in Elizabeth City, NC experiences frequent flood conditions seemingly unrelated to rainfall amounts. The purpose of this study was to compare barometric pressures, lunar cycles, wind directions, and wind speeds with water depth readings of Charles Creek (a tributary of the Pasquotank River). A static remote imaging system was used to measure water depth through pixel enumeration and referencing through remote sensing techniques coupled with custom image processing software. Environmental data was collected through Elizabeth City State University's National Renewable Energy Laboratory Weather Monitoring Station. The data was correlated using MINITAB (R) to find an equation to approximate a model of the rise of the Creek water level based upon environmental conditions. The regression equation had a coefficient of determination of 42%; this means there was a 42% probability the model was useful at predicting the pixel count based on environmental variables.**

*Keywords-Floods; Atmospheric Modeling; Image Processing; Remote Sensing; Regression Analysis*

## I. INTRODUCTION

The Elizabeth City Watershed in North Carolina serves as a tributary to the Pasquotank River and has long puzzled town officials due to its flood tendencies seemingly unrelated to rainfall [1]. Residents assumed the upstream flooding was caused by rapid industrial development in the watershed as well as increased sediment deposits, additional clogging from downed trees, and clogged culverts being the primary issues. Using computer modeling and on site observations, a Norfolk, Virginian, named Greg Johnson (an engineer and director of engineering for Patton Harris Rust and Associates), studied the heavy flooding in areas within the Knobbs Creek drainage district of Pasquotank County [1]. Johnson's function was to pinpoint both the cause of flooding and designate which areas were likely to flood. Recommendations included additional canals, additional ditches, dredging of debris from swamps, and requiring developers of new projects to include increased storm water detention ponds [2]. Unfortunately, in August and November of 2009, storms hit the area-which delayed and eventually discontinued the study [9].

The research study of Charles Creek flood zone area will result in the creation of a model to correlate barometric pressure, lunar cycles, wind speed, and wind direction on the water levels in the creek. During the study several general questions will be answered associated with a predicted cause of the flooding. The questions related to the cause of flooding will state if barometric pressure has a correlative relationship with the Charles Creek water level? Can a visual method be used to capture images in order to measure water levels in Charles Creek? Which environmental variable has a greater effect on the water level? Lastly, what equation can model the interaction between water level and the environmental variables? The team believed barometric pressure and lunar cycles would have the largest effect on the model of the Charles Creek water level.

## II. NATURE OF STUDY

Environmental conditions thought to influence flood conditions include: barometric pressure, wind speed, and wind direction [9], lunar cycles, and sea level [9]. Barometric pressure, or atmospheric pressure, is a measure of how much pressure is exerted on an area. Measured in millibars, the metric units of air pressure, barometric pressure is a function of how many molecules of air there are in a specific area. Since air pressure can change rapidly with different altitudes, a standard is to adjust all locations to sea level before a comparison can be concluded [3]. Using isobaric maps and connecting points of equal pressure, the strength and direction of the wind can be determined [4]. Furthermore, a lunar phase of the moon as seen by an observer on Earth can have an impact on environmental conditions, such as the sea level. Mean sea level is a measure of the average height of the

ocean's surface with respect to the land. When sea level rises and falls due to the attraction of the moon, tides are produced. The sun and moon are aligned, which cause a strong gravitational pull and extremely high or low tides; these tides are referred to spring tides and occur during the full and new moon phases. On the other hand, neap tides occur during the moon's quarter phases. The tides are weaker and occur when the gravitational forces of the moon and sun are perpendicular to one another with respect to the Earth [6].

In particular, the National Renewable Energy Laboratory (NREL), previously named the Solar Energy Research Lab (SERI) [18], maintains the Renewable Resource Data Center and allows users to access a collection of renewable energy resources, such as data maps and tools. Elizabeth City State University installed a weather station in 1985 [18] in conjunction with NREL to provide data sets from past decades including: daily plots, raw data files, and live displays [6]. It is

located at a latitude of 36.28 ° North and a longitude of 76.22 ° West at an elevation of 26 meters above mean sea level on the Jenkins Science Center. The station located atop the roof of the Jenkins Science Center is pictured in Figure 1.

The moon's phase is measured as an illumination percentage. To receive the percentage of the illuminated moon, one can visit the Brian Casey Artifact; a website which allows a user to enter a specific date to find out the percentage of illumination of the moon based on midnight United States Central time as viewed from the Northern Hemisphere [8].

### A. Modeling

In the 1950's, researchers began developing the first computer models for the Earth's atmosphere [10]. The computer models used mathematical equations, which could predict the physical changes on weather for years, or even centuries in advance. The first models were rudimentary, but provided valuable insight into the atmosphere's future state. The models were never intended to replace human forecast, and the atmospheric equations were too complex forthen modern computers. Human forecasters continued to study the results from models over periods of time to compare forecast output from the models to actual weather occurrences.

Many new weather models are currently being developed [11]. The different types of models vary between climate, mesoscale, dynamic, and statistical models. Climate models are primarily used to forecast changes in the earth's climate. Mesoscale models are mainly used to forecast the weather. Dynamic models are the most sophisticated and costly tools used to forecast the weather and use fundamental equations of the atmosphere to predict changes in the weather based on current conditions. Lastly, statistical models are primarily used to help meteorologist track weather systems [12].

The FLO-2D model is a volume conservation, two-dimensional flood routing model that distribute a flood hydro graph over a system of square grid elements using a finite difference algorithm. Am experiment used the model to create virtual tsunami waves and ocean storm surges with FLO-2D [19]. FLO-2D effectively analyzes riverine or unconfined fan flooding. J.S. O'Brien's research came from the city of



Figure 1

Waikiki, Oahu, Hawaii using an existing FLO-2D watershed model developed by Hawaii District Corps of Engineers. The goal of the Corp's investigation was to describe the flood hazard for various return period flood events and then using those results of flood hazard maps to quantify economic benefits to justify federal participation in the construction of an Ala Wai flood damage reduction project. With minimal information consisting of a digital terrain model and ocean water surface as a function, the FLO-2D model was used to generate detailed mapping associated with ocean storm surges or tsunamis. The research also produced an available wave height model to create the time stage data pairs, therefore using the FLO-2D model to predict coastal flood hazard areas associated with storm surge or tsunamis worldwide.

### B. Remote Sensing

When using remote sensing techniques, it is common practice to qualify a sensor by finding field of view and resolution [11]. Field of view is the area or solid angle, which can be viewed through an instrument using eyesight [10]. It is also defined as the angle in which objects are recorded on the film or sensor in a camera [13]. Resolution is the number of pixels contained on a display monitor, expressed in terms of the number of pixels on the horizontal axis and the number on the vertical axis [12]. It is also the term used to describe the number of dots and pixels to display an image [12].

An image processing software, Environment for Visualizing Images ® (ENVI), can be used to count the number pixels of a particular object in an image and as a means to measure the sea level daily [14]. It is an extremely powerful tool for analyzing the content of hyper-spectral images. It combines complete processing with advanced spectral tools [15]. Some other functions of the software application are image enhancement and feature extraction. ENVI ® is tightly integrated with ArcGIS (R) from Esri.

## C. Statistical Analysis

Correlation is the relationship between two or more variables. In terms of defining correlations there are two basic types called negative and positive correlation. Negative correlation is the relationship between two variables while in opposite directions [21]. Positive correlation is the relationship between two variables related in similar ways. This means if one decreases or increases the other does the same and vise verse [22]. Regression is a statistical attempt to determine the relationship strength between one dependent variable and a series of other changing variables. Regression also takes a group of random variables, and tries to find a mathematical relationship between them, which is usually in the form of a straight line [23]. A regression equation has many variables such as $R^2$, P-Value, T-Value, Correlation Coefficients, Predictors, and constants. $R^2$ is most often called the coefficient of determination, which is defined as the ratio of the sum of squares using a regression model [24]. Its main purpose is to predict future outcomes on other related information within a data set, which is mostly used in a linear regression [25]. P-Values of the equation are mostly based off of assumptions in statistical hypothesis testing. It indicates the possibility the result obtained in statistical test is due to a probability rather than a true relationship between measures [26]. The t-value is a hypothesis test used in comparing two groups of data whenever the null hypothesis is not true while analyzing data [27]. Correlation coefficients measures determine the degree to which two variable movements are associated [29]. It also measures how well predicted values fit into real life data referred to straight line or linear relationships between two variables [28]. Predictors are significantly independent variables to predict dependent variables in regression analysis [26].

## III. METHODOLOGY

### A. Materials

Most of the hardware involved in data collection was not complex compared to the advanced software used in data processing. The remote imaging system was installed inside a local resident's home. The install location was befitting in that the creek was directly in the backyard. Positioned in the window, the web camera model used to capture images was a Creative Live Cam Socialize HD (VF0610) with a resolution of 5.7 Mega Pixels, focal length of 10 mm, width of field of 19.89°, and a height of field of 14.97°. It was attached to an ASUS EEE personal computer. The laptop had a 1.6 GHz processor, 1GB of RAM, and 160 GB of hard drive space. With the Windows 7 professional operating system installed, the laptop was connected to power and used both the internal battery as well as a Tripplite® Battery Backup for backup power. The battery and laptxop are shown in Figure 2.



**Figure 2**

The 1.0.7 version of ChronoLapse was used as the software to take time-lapse images to be viewed later in ENVI®, a specialized image processing software. A seven-foot aluminum-fencing pole was obtained and painted with fluorescent pink spray paint so the object would contrast against colors in nature as shown in Figure 3.



**Figure 3**

Dropbox® was used as a cloud based image storage for all images taken as a backup to flash drives. Additionally, Google® Docs was used to store all documents, facilitating the use and revision of all reports by team members and mentors. When fieldwork was conducted, a standard tape measurer including metric units of measurement was used as a tool to measure the visible height of the pole. Most importantly, the Minitab statistical software served as a means to create a linear regression of the data.

### B. Delimiters

Factors eliminated from research and data collection include:

- Pressure gauges (barometer / tide gauges)

- The contribution of rainfall and/or runoff

- Hydraulic Models

- Night Captures (Figure 4)

- Glints from sunlight in images (Figure 5)


**Figure 4**


**Figure 5**

*C. Setup*

The team installed the remote imaging system and other materials on the morning of June 8, 2011 at 9:24 am. The pole was driven into the creek approximately two feet deep, initially with the top of the pole elevation reading -7 feet (213.36 centimeters) at GPS location North 36 17.678 and West 076 13.042, which was 1024.8 centimeters away from the camera location. The first reading of the visible pole (from the top of the exposed pole down to the Charles Creek water level) read 124.7775 centimeters. As for the remote imaging setup, Camorama was installed and set to capture on a thirty-minute interval with a 640 X 360 resolution. Camorama is an unsupported open source software application in which we ran on an Ubuntu platform. The sleep mode and automatic update feature on the laptop was disabled due to its potentiality to interfere with the program.

*D. Procedures*

Upon arriving onsite to collect images the first task was to clean the optics to ensure the images would continue to be clear. After doing so the pictures were saved to a removable disk and the laptop was checked for any malfunctions. A few force captures were also made, causing the times to not remain on the same 30-minute interval. To load the file names for the pictures onto the Google® Docs spreadsheet a bash script was

created by Jerome Mitchell (Indiana University). To execute the script, [. /epoch. sh] was typed into the command line of a UNIX terminal. The script took the UNIX Time Code (UTC) timestamp from the photos to give the time and date of the picture. After copying the file names temporarily into a Microsoft Excel spreadsheet for sorting, they were then transferred into Google® Docs. (Prior to the creation of the script, adding the image information was done manually through the Epoch Converter website.)

To obtain the height in pixels of the pole pixel enumeration was conducted using ENVI®. By selecting the proper folder containing the pictures in Dropbox, it allowed the team to load RGB images into the program. The measurement tool was used on the zoomed in image to select the top of the pole and lowest visible point on the pole to obtain the actual height of the visible poleto four significant digits. From this point, the team could estimate the height of the sea level. This information was then added to the Google® Docs spreadsheet.

Monitored independent variables, for environmental conditions were collected online either using the Elizabeth City State University Weather Station or the Brian Casey Artifact Page. On the ECSU weather station site, the information was found within the "Daily Plots and Raw Data Files" under the time frame beginning with April 1, 2004 to the current date. After selecting the desired date along with station pressure, wind speed and, wind direction as the 5-minute interval ASCII text, all information was loaded for recording into a spreadsheet in Google® Docs. Pictures captured after 7:20 pm and before 4:50 am were clear enough to conduct pixel enumeration were assigned hourly data readings due to limitations of the ECSU website. The Brian Casey Artifact website only required input dates to present the illumination percentage of the moon as it related to lunar phases. All information was then added into the master Google® Docs spreadsheet.

IV.  RESULTS

In the beginning of data collection, it was noted a few of the images were light at times of the day when it should have been dark outside. The program Camorama, showed an error message but the file names were all unique. Therefore, the same picture would be used for several different files with duplicate images. On June 20, 2011, the software was changed from Camorama to ChronoLapse after testing the program in the lab by capturing images every 15 seconds for a period of 90 minutes. A Google® supported code, ChronoLapse, is a tool used for capturing photos in series on Windows operating system using screen captures [17]. It is capable of taking screenshots or webcam captures at user-defined intervals [17]. Each piece of the software's functionality was created to work along with making the program more versatile [17]. Other features to come with this software includes: dual monitor support and picture in picture [16].The operating system platform was also changed from Ubuntu to Windows 7 Professional. As a result, an abundance of data was lost from the time between setup and discovering the error.

To facilitate the processing of the images, the team initially attempted to create an automated image processing algorithm using MATHWORKS MATLAB ®. The focus was to create a system, which would automatically recognize the pole in the images based on pixel colors and the location in the image. The drawback to the system was the MATLAB ® Image Processing Toolbox was unavailable at the time of research. In the same fashion, the MATLAB® Statistical Toolbox was unavailable to perform a multipleregression, which resulted in the use of Minitab®, application software.

The regression equation derived to model the water level of the creek based on the recorded environmental conditions is listed in Figure 6

**Equation 1**

Pixels= 33.5- 0.0067* (Barometric Pressure in mBars)

-0.00194* (Wind Direction degrees from North)

-0.122* (Wind Speed in m/s)

+4.29 * (Lunar Cycle Illuminated Fraction Percentage)

Based on the given $R^2$ value there was 42% coefficient of determination explained by the four different independent variables, meaning there was a 42% probability the model was useful at predicting the pixel count based on given values for barometric pressure, lunar cycles, wind speed, and wind direction. The chart with the trend line is show in Figure 6.
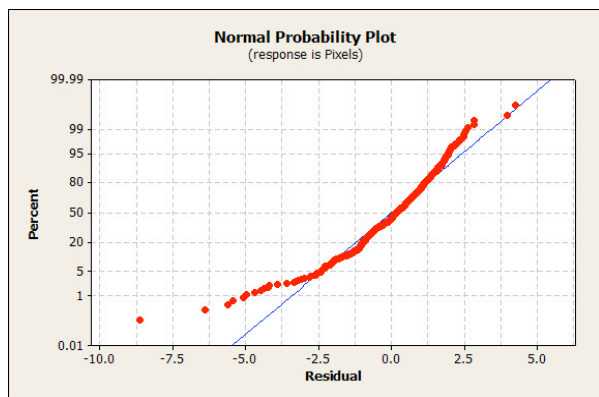

**Figure 6**

Lunar Cycles, alone had the greatest contribution to the model with a $R^2$ value of 40.5%. The regression model plotted with outliers is shown in Figure 7.
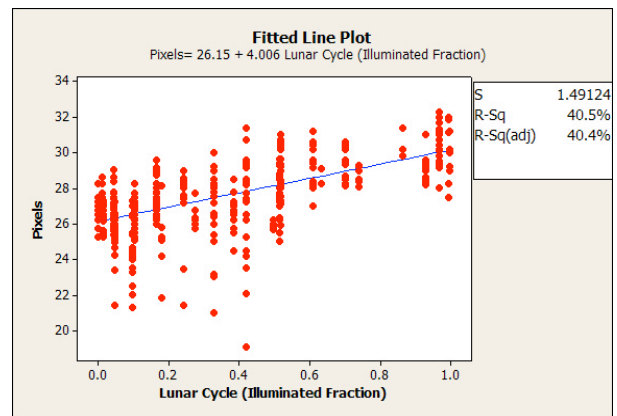

**Figure 7**

Lunar Cycles and wind direction jointly had the second highest variance of 41.5%. Lastly, the two other independent variables with a significant prediction value were Lunar Cycles and Wind Speed at 41.2%.

## V.    CONCLUSION

In conclusion, the flood zone modeling project purpose was to find the cause of flooding in Charles creek in Elizabeth City, North Carolina. The procedure to do so was to create a model to test each environmental variable individually and combined to see which variable had the greatest effect on the rise of water level in Charles Creek. Throughout the project the data analysis and step-by-step procedure presented answers to the sub-questions throughout the research paper. The barometric pressure placed inside of the model creation by the flood zone modeling team showed it had very little correlative relation with the Charles Creek water level. Barometric Pressure only had a coefficient of determination value of 2.5% aside from other predictor variables. Furthermore, the data analysis did show the lunar cycle variable, individually, placed in the model showed greater effect on the water level of Charles Creek than others with a coefficient of determination of 40.5%. The $R^2$ value for wind direction was 0% and 2.1% for wind direction. Although wind direction and wind speed had some effect on the total prediction capability of the model, it was not enough to state they were sufficient predictors. Ultimately, the environmental variables placed inside the model together showed the greatest effect on the Charles Creek water level at a 42% probability. The equation modeled the interaction between the water level and environmental variables are shown in Equation 1. Finally, visual methods proved to be helpful in capturing images in order to messure water levels in Charles Creek and facilitated the team's procedures and derive a meaningful model from the data analysis.

## VI.    FUTURE WORK

The future goal of this project is to create a program to facilitate automated image processing in MATLAB ® or another high level programming language. By doing so, it will speed up the time it takes to get the pixel count in photos as well as provide a more accurate count of the pixels. Another advantage to having the program to run on the files means a

greater number of images can be processed. Therefore, instead of having the remote imaging system take pictures on a 30-minute interval, it could be increased to match weather station measurements of five-minute intervals. As a result, this would improve the data set and increase the prediction capability of the model. A lighting system could also be used to illuminate the pole so pictures could be captured at night.

The use of a cloud-based image storing system with an Internet connection at the site would eliminate much of the fieldwork of manually saving photos. Altogether for some variation, in the future, a pressure gauge could be used as a comparison to pixel enumeration to validate the imaging method.

## References

[1] J. Hampton, "Flood-Prone Elizabeth City Neighborhood's a Puzzle," *The Virginian-Pilot*, Norfolk, January 16, 2009.

[2] E. C. D. A. (2011, June 9) Action on Bridge Should Relieve Flooding. [Online]. Available: http://www.dailyadvance.com/uncategorized/action-bridge-should-relieve-flooding-64614

[3] L. Adams. (2011, June 16) Effects of Barometric Pressure. [Online]. Available: http://www.quickoneplus.com/fish/articles/page.asppage=barometric

[4] W. O. U. (2011, June 16) Weather Facts: Isobars on Surface Maps. [Online]. Available: http://www.weatheronline.co.uk/reports/wxfacts/Isobars-on-surface-maps.htm

[5] W. (2011, June 25) Glossary of Weather Terms. [Online]. Available: http://weather.weatherbug.com/weather-glossary/weather-glossary.html

[6] K. Cooley. (2011, June 26) The Moon and Tides: How The Moon Affects Ocean Tides. [Online]. Available: http://home.hiwaay.net/~krcool/Astro/moon/moontides/

[7] N. R. E. L. (2011, June 26) NREL: MIDC/Elizabeth City State University. [Online]. Available: http://www.nrel.gov/midc/ecsu/

[8] B. Casey. (2011, June 26) The Moon's Phase. [Online]. Available: http://www.briancasey.org/artifacts/astro/moon.cgi

[9] G. Johnson, "Interview," Phone,

[10] R. Atkins. (2011, June 26) Field of View. [Online]. Available: http://photo.net/learn/fov/

[11] (2011, June 26) Field of View. [Online]. Available: http://www.answers.com/topic/field-of-view

[12] W. Russell. (2011, June 27) Resolution. [Online]. Available: http://presentationsoft.about.com/od/r/g/resolution.htm

[13] (2011, June 22) Resolution . [Online]. Available: http://searchcio-midmarket.techtarget.com/definition/resolution

[14] (2011,June 22)ENVI. [Online]. Available: http://acronyms.thefreedictionary.com/ENVI

[15] A. R. S. C. S. (2011, June 23) ENVI. [Online]. Available: http://www.arsc.edu/support/resources/envi.html

[16] (2011, June 20) ChronoLapse. [Online]. Available: http://code.google.com/p/chronolapse/

[17] C. Green. (2011, June 20) ChronoLapse. [Online]. Available: http://keeyai.com/projects-and-releases/chronolapse/

[18] S. Khan, "ECSU Weather Station," electronic mail,

[19] J. O'Brien. (2011, June 26) MODELING TSUNAMI WAVES AND OCEAN STORM SURGES WITH FLO-2. [Online]. Available: http://www.flo-2d.com/wp-content/uploads/tsunami-and-storm-surge.pdf

[20] Hollowell and Hoover. (2011, June 9) Pasquotank River Watershed Integrated Water Design (IWD) Project. [Online]. Available: http://www.soil.ncsu.edu/about/news/Pasquotank.html

[21] E. Tuomey. (2011, July 2) What Does 'Correlation' Mean? [Online]. Available: http://www.writefix.com/graphs/correlation.htm

[22] (2011, July 2) Positive Correlation. [Online]. Available: http://www.investopedia.com/terms/p/positive-correlation.asp

[23] (2011, July 2) Regression. [Online]. Available: http://www.investopedia.com/terms/r/regression.asp

[24] N. Henry. (2011, July 2) R-square and Standardization in Regression . [Online]. Available: http://www.people.vcu.edu/~nhenry/Rsq.htm

[25] (2011, July 3) R-Squared. [Online]. Available: http://www.trade-ideas.com/Glossary/R-Squared.html

[26] M. Mutchler. (2011, July 3) Glossary of Statistical Terms. [Online]. Available: http://www.apla.org/research/REC/glossary.htm

[27] W. Trochim. (2011, July 7) The T-Test. [Online]. Available: http://www.socialresearchmethods.net/kb/stat_t.php

[28] (2011, July 7) What Is the Correlation Coefficient. [Online]. Available: http://www.forecasts.org/cc.htm

[29] (2011, July 7) Correlation. [Online]. Available: ]http://www.investopedia.com/terms/c/correlationcoefficient.asp